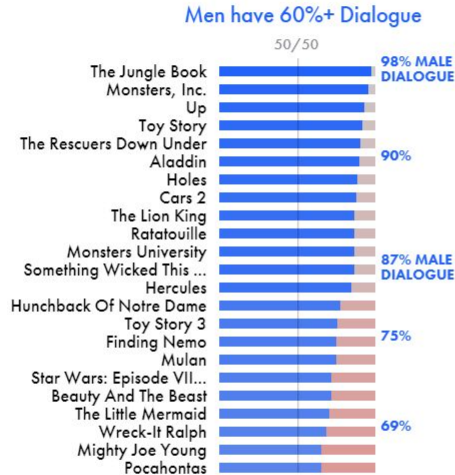# Hollywood sticks to the script: Films aren't more inclusive, despite a decade of advocacy
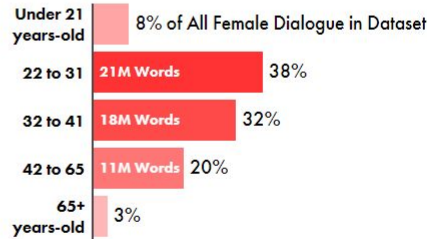
Quantification of Gender Representation Bias in Commercial Films based on Image Analysis
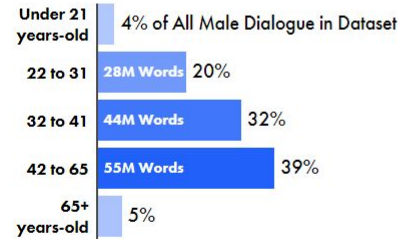
Disney Screenplay Dialogue, Broken-down by Gender

Men have 60%+ Dialogue

50/50

Film	
The Jungle Book	98% MALE DIALOGUE
Monsters, Inc.	
Up	
Toy Story	
The Rescuers Down Under	90%
Aladdin	
Holes	
Cars 2	
The Lion King	
Ratatouille	
Monsters University	87% MALE DIALOGUE
Something Wicked This ...	
Hercules	
Hunchback Of Notre Dame	
Toy Story 3	
Finding Nemo	75%
Mulan	
Star Wars: Episode VII...	
Beauty And The Beast	
The Little Mermaid	
Wreck-It Ralph	69%
Mighty Joe Young	
Pocahontas	

Look Who's Still Talking the Most in Movies: White Men

Actresses: Total Words, by Age

Age	Words	%
Under 21 years-old		8% of All Female Dialogue in Dataset
22 to 31	21M Words	38%
32 to 41	18M Words	32%
42 to 65	11M Words	20%
65+ years-old		3%

Actors: Total Words, by Age

Age	Words	%
Under 21 years-old		4% of All Male Dialogue in Dataset
22 to 31	28M Words	20%
32 to 41	44M Words	32%
42 to 65	55M Words	39%
65+ years-old		5%

Adriana Rakshana
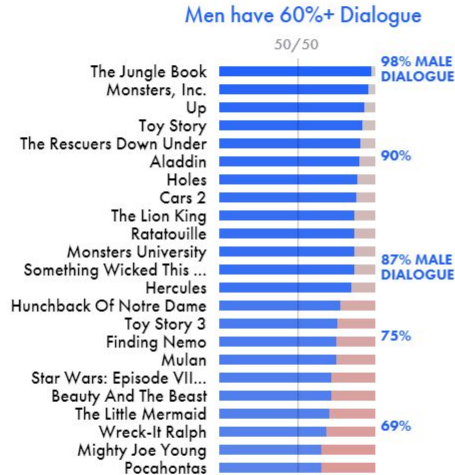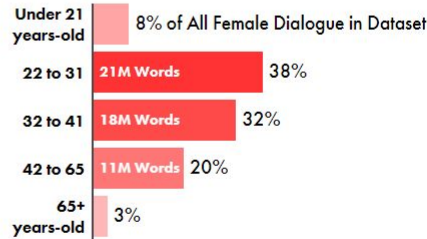
Hollywood sticks to the script: Films aren't more inclusive, despite a decade of advocacy

Quantification of Gender Representation Bias in Commercial Films based on Image Analysis

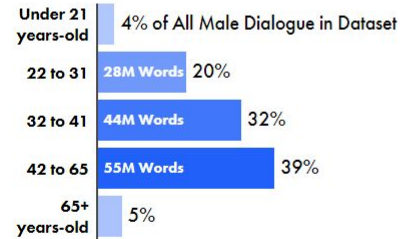### Screenplay Dialogue, Broken-down by Gender

Disney

Men have 60%+ Dialogue

50/50

Film	
The Jungle Book	98% MALE DIALOGUE
Monsters, Inc.	
Up	
Toy Story	
The Rescuers Down Under	90%
Aladdin	
Holes	
Cars 2	
The Lion King	
Ratatouille	
Monsters University	87% MALE DIALOGUE
Something Wicked This ...	
Hercules	
Hunchback Of Notre Dame	
Toy Story 3	
Finding Nemo	75%
Mulan	
Star Wars: Episode VII...	
Beauty And The Beast	
The Little Mermaid	
Wreck-It Ralph	69%
Mighty Joe Young	
Pocahontas	

Look Who's Still Talking the Most in Movies: White Men

Actresses: Total Words, by Age

Age	Words	%
Under 21 years-old		8% of All Female Dialogue in Dataset
22 to 31	21M Words	38%
32 to 41	18M Words	32%
42 to 65	11M Words	20%
65+ years-old		3%

Actors: Total Words, by Age

Age	Words	%
Under 21 years-old		4% of All Male Dialogue in Dataset
22 to 31	28M Words	20%
32 to 41	44M Words	32%
42 to 65	55M Words	39%
65+ years-old		5%

Adriana Rakshana

Data and Methodology

Around 10000 movies from the Davies corpus

Sorted by decades from 1930s to the 2010s ending in 2017

Separated by genres

- Vector Space Word Embeddings
- Word occurrence and Frequencies
- Semantic Networks
- Parts of Speech (NV and NJ)

Data and Methodology

Around 10000 movies from the Davies corpus

Sorted by decades from 1930s to the 2010s ending in 2017

Separated by genres

- Vector Space Word Embeddings
- Word occurrence and Frequencies
- Semantic Networks
- Parts of Speech (NV and NJ)

Data and Methodology

Around 10000 movies from the Davies corpus

Sorted by decades from 1930s to the 2010s ending in 2017

Separated by genres

- Vector Space Word Embeddings
- Word occurrence and Frequencies
- Semantic Networks
- Parts of Speech (NV and NJ)

Data and Methodology

Around 10000 movies from the Davies corpus

Sorted by decades from 1930s to the 2010s ending in 2017

Separated by genres

- Vector Space Word Embeddings
- Word occurrence and Frequencies
- Semantic Networks
- Parts of Speech (NV and NJ)

Main Insight

Time Related Word Evolution is more revealing than Genre Comparisons

Word similarities for men and women are very different

Gender stereotypes have evolved?

```
mvD2V_female.most_similar(positive = ['woman','girl'], negative = ['man'], topn = 10)
```

```
[('brunette', 0.5521952509880066),
 ('blonde', 0.547403872013092),
 ('redhead', 0.5351743698120117),
 ('prettiest', 0.5166902542114258),
 ('pinup', 0.5048144459724426),
 ('flirt', 0.4958895742893219),
 ('minder', 0.4949256181716919),
 ('niece', 0.489028662443161),
 ('bombshells', 0.48542696237564087),
 ('blond', 0.48342978954315186)]
```

Tagged Female

```
mvD2V_male.most_similar(positive = ['woman','girl'], negative = ['man'], topn = 10)
```

```
[('redhead', 0.5722601413726807),
 ('brunette', 0.5712414979934692),
 ('blonde', 0.565335750579834),
 ('mousy', 0.51931190490072266),
 ('unisex', 0.5185565948486328),
 ('bombshells', 0.5135267972946167),
 ('prettiest', 0.5078239440917969),
 ('daughter', 0.49182310700416565),
 ('minx', 0.4886482357978821),
 ('carla', 0.48468464612960815)]
```

Tagged Male

W
O
M
E
N

```
mvD2V_female.most_similar(positive = ['woman','girl'], negative = ['man'], topn = 10)
```

```
[('brunette', 0.5521952509880066),
 ('blonde', 0.547403872013092),
 ('redhead', 0.5351743698120117),
 ('prettiest', 0.5166902542114258),
 ('pinup', 0.5048144459724426),
 ('flirt', 0.4958895742893219),
 ('minder', 0.4949256181716919),
 ('niece', 0.489028662443161),
 ('bombshells', 0.48542696237564087),
 ('blond', 0.48342978954315186)]
```

Tagged Female

```
mvD2V_male.most_similar(positive = ['woman','girl'], negative = ['man'], topn = 10)
```

```
[('redhead', 0.5722601413726807),
 ('brunette', 0.5712414979934692),
 ('blonde', 0.565335750579834),
 ('mousy', 0.51931190049072266),
 ('unisex', 0.5185565948486328),
 ('bombshells', 0.5135267972946167),
 ('prettiest', 0.5078239440917969),
 ('daughter', 0.49182310700416565),
 ('minx', 0.4886482357978821),
 ('carla', 0.48468464612960815)]
```

Tagged Male

W
O
M
E
N

```
mvD2V_female.most_similar(positive = ['man','boy'], negative = ['woman'], topn = 10)
```

```
[('buddy', 0.5500340461730957),
 ('rascals', 0.5384502410888672),
 ('homie', 0.5353148579597473),
 ('dawg', 0.5141319036483765),
 ('lad', 0.502248227596283),
 ('homies', 0.5002907514572144),
 ('marky', 0.4982272684574127),
 ('homey', 0.4966511130332947),
 ('fam', 0.49355340003967285),
 ('homeboy', 0.49234485626220703)]
```

```
mvD2V_male.most_similar(positive = ['man','boy'], negative = ['woman'], topn = 10)
```

```
[('buddy', 0.5860825181007385),
 ('atta', 0.5367481708526611),
 ('jangers', 0.5269747972488403),
 ('jamesy', 0.5155718922615051),
 ('lad', 0.5068671703338623),
 ('homie', 0.5006662607192993),
 ('scout', 0.48949190974235535),
 ('dink', 0.4848521053791046),
 ('bro', 0.48441779613494873),
 ('drummer', 0.4813867509365082)]
```

```
mvD2V_female50.most_similar(positive = ['woman','girl'], negative = ['boy', 'man'], topn = 10)
```

```
[('faithful', 0.8399957418441772),
 ('sophisticated', 0.8392194509506226),
 ('attractive', 0.8286294937133789),
 ('mustache', 0.8201954364776611),
 ('unreal', 0.8033789396286011),
 ('marriage', 0.7966414093971252),
 ('quality', 0.7872207164764404),
 ('modest', 0.7836649417877197),
 ('farthing', 0.7808701992034912),
 ('women', 0.7785411477088928)]
```

1950s

```
mvD2V_female10.most_similar(positive = ['woman','girl'], negative = ['boy', 'man'], topn = 10)
```

```
[('attractive', 0.5501745939254761),
 ('lady', 0.5045493841171265),
 ('sexy', 0.45774829387664795),
 ('women', 0.45704615116119385),
 ('virtuous', 0.4526621103286743),
 ('lesbian', 0.4518219232559204),
 ('vagina', 0.43385642766952515),
 ('succubus', 0.43176722526550293),
 ('extensions', 0.43111667037010193),
 ('passionately', 0.4297320246696472)]
```

2010s

```
mvD2V_female50.most_similar(positive = ['woman','girl'], negative = ['boy', 'man'], topn = 10)
```

```
[('faithful', 0.8399957418441772),
 ('sophisticated', 0.8392194509506226),
 ('attractive', 0.8286294937133789),
 ('mustache', 0.8201954364776611),
 ('unreal', 0.8033789396286011),
 ('marriage', 0.7966414093971252),
 ('quality', 0.7872207164764404),
 ('modest', 0.7836649417877197),
 ('farthing', 0.7808701992034912),
 ('women', 0.7785411477088928)]
```

1950s

```
mvD2V_female10.most_similar(positive = ['woman','girl'], negative = ['boy', 'man'], topn = 10)
```

```
[('attractive', 0.5501745939254761),
 ('lady', 0.5045493841171265),
 ('sexy', 0.45774829387664795),
 ('women', 0.45704615116119385),
 ('virtuous', 0.4526621103286743),
 ('lesbian', 0.4518219232559204),
 ('vagina', 0.43385642766952515),
 ('succubus', 0.43176722526550293),
 ('extensions', 0.43111667037010193),
 ('passionately', 0.4297320246696472)]
```

2010s

```
mvD2V_female50.most_similar(positive = ['woman','girl'], negative = ['boy', 'man'], topn = 10)
```

```
[('faithful', 0.8399957418441772),
 ('sophisticated', 0.8392194509506226),
 ('attractive', 0.8286294937133789),
 ('mustache', 0.8201954364776611),
 ('unreal', 0.8033789396286011),
 ('marriage', 0.7966414093971252),
 ('quality', 0.7872207164764404),
 ('modest', 0.7836649417877197),
 ('farthing', 0.7808701992034912),
 ('women', 0.7785411477088928)]
```

1950s

```
mvD2V_female10.most_similar(positive = ['woman','girl'], negative = ['boy', 'man'], topn = 10)
```

```
[('attractive', 0.5501745939254761),
 ('lady', 0.5045493841171265),
 ('sexy', 0.45774829387664795),
 ('women', 0.45704615116119385),
 ('virtuous', 0.4526621103286743),
 ('lesbian', 0.4518219232559204),
 ('vagina', 0.43385642766952515),
 ('succubus', 0.43176722526550293),
 ('extensions', 0.43111667037010193),
 ('passionately', 0.4297320246696472)]
```

2010s

```
mvD2V_female30.most_similar(positive = ['woman','girl'], negative = ['boy','man'], topn = 10)
```

```
[('danced', 0.8751518726348877),
 ('wonderful', 0.8557447195053101),
 ('grown', 0.8504206538200378),
 ('happy', 0.8413878679275513),
 ('beautiful', 0.8326112031936646),
 ('husbands', 0.8314074873924255),
 ('married', 0.8272198438644409),
 ('grandest', 0.8158673048019409),
 ('witty', 0.8075016736984253),
 ('cared', 0.8072414398193359)]
```

1930s

```
mvD2V_female00.most_similar(positive = ['woman','girl'], negative = ['boy', 'man'], topn = 10)
```

```
[('ayoung', 0.421100914478302),
 ('romantic', 0.4095713794231415),
 ('exhibitionist', 0.4050207734107971),
 ('monogamous', 0.4032568633556366),
 ('attractive', 0.40316757559776306),
 ('flirts', 0.398554265499115),
 ('reesa', 0.3958331048488617),
 ('sex', 0.3800505995750427),
 ('husband', 0.3777911961078644),
 ('tantric', 0.3761769235134125)]
```

2000s

```
mvD2V_rom_male.most_similar(positive = ['woman','girl'], negative = ['man'], topn = 10)
```

```
[('prettiest', 0.5798849123954773),
 ('goldenrod', 0.5258733034133911),
 ('blond', 0.5102574825286865),
 ('attractive', 0.4982985258102417),
 ('redhead', 0.4951922297477722),
 ('insanely', 0.4912869334220886),
 ('mermaid', 0.4798300266265869),
 ('wilma', 0.4768679141998291),
 ('adele', 0.47348636388778687),
 ('paraded', 0.47293227910995483)]
```

```
mvD2V_com_female.most_similar(positive = ['woman','girl'], negative = ['man', 'boy'], topn = 10)
```

```
[('redhead', 0.5138537883758545),
 ('peignoir', 0.5102139711380005),
 ('prettiest', 0.5037169456481934),
 ('attractive', 0.5025854110717773),
 ('scrubba', 0.492868572473526),
 ('charmer', 0.4899721145629883),
 ('blonde', 0.47883540391921997),
 ('flirt', 0.46782612800598145),
 ('daughter', 0.46521151065826416),
 ('schoolgirl', 0.4647250175476074)]
```

SPORT
```
[('cynthia', 0.8991438150405884),
 ('cheeky', 0.883186399936676),
 ('mood', 0.8707400560379028),
 ('diving', 0.870449423789978),
 ('wife', 0.8673449158668518),
 ('lovers', 0.8655206561088562),
 ('sticky', 0.8611229658126831),
 ('stig', 0.85368412733078),
 ('tattoo', 0.8490241765975952),
 ('drawer', 0.8488390445709229)]
```
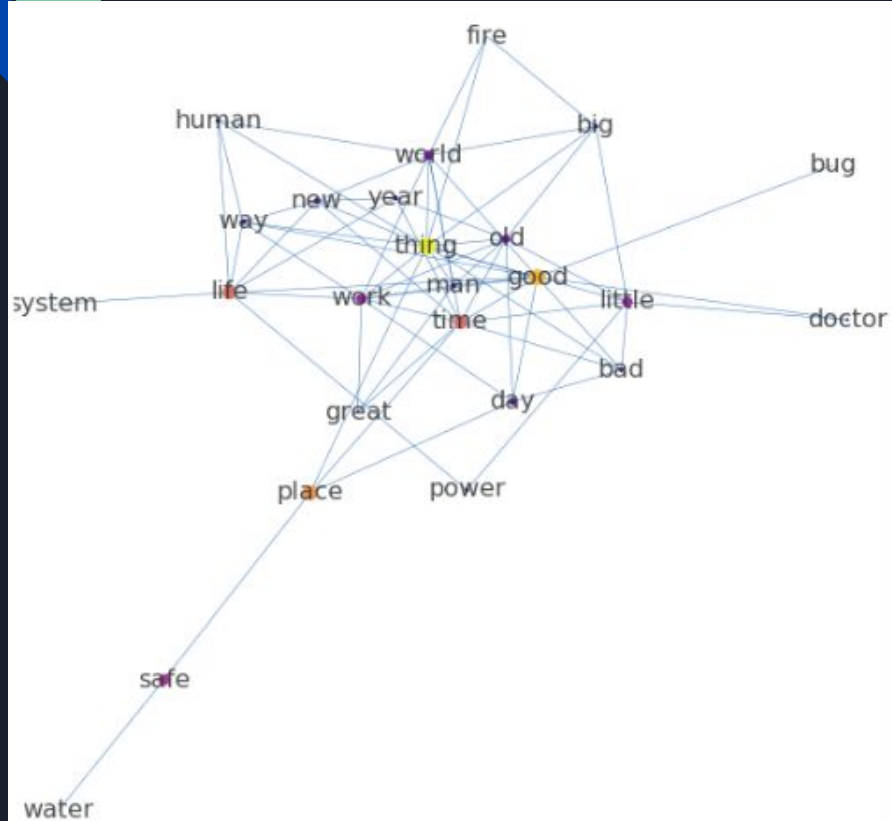
SCI-FI

HORROR
```
[('beezel', 0.5858185887336731),
 ('child', 0.5797515511512756),
 ('prettiest', 0.5778923034667969),
 ('sister', 0.5757694840431213),
 ('orphan', 0.5677846670150757),
 ('judie', 0.5676324367523193),
 ('terrace', 0.5632544159889221),
 ('diary', 0.5605330467224121),
 ('mancini', 0.5568317174911499),
 ('angel', 0.5518367290496826)]
```
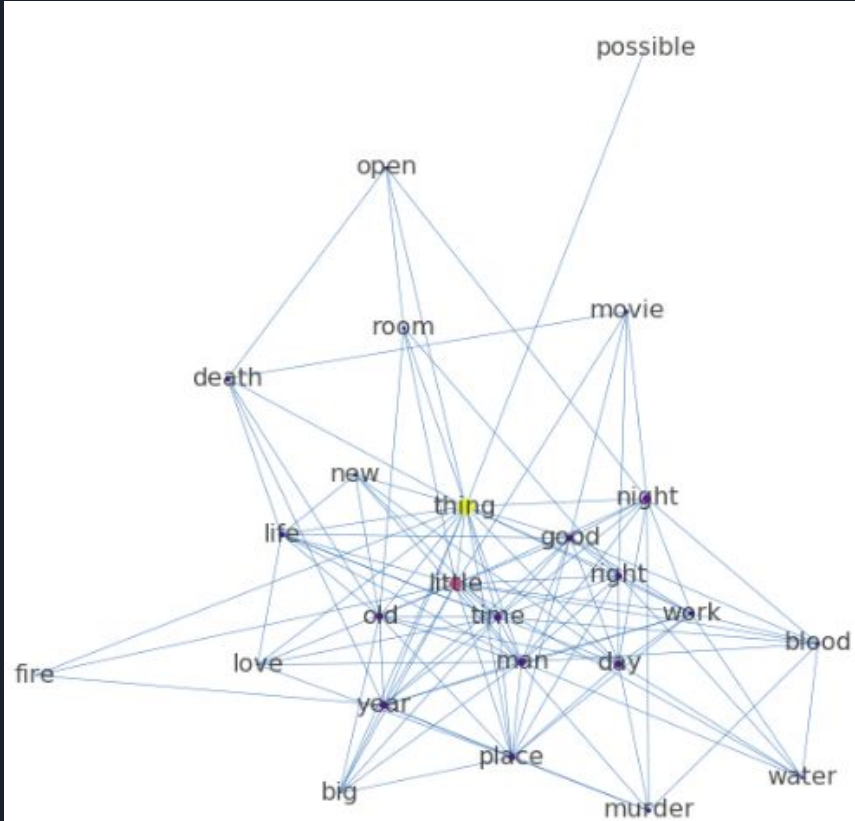
```
[('flower', 0.7666660054725647),
 ('attractive', 0.750607967376709),
 ('wookiee', 0.7502974271774292),
 ('brunette', 0.7406696081161499),
 ('sister', 0.7392488718032837),
 ('poor', 0.7329496145248413),
 ('christmas', 0.7321585416793823),
 ('sweetest', 0.7285386323928833),
 ('brighten', 0.7274667024612427),
 ('loved', 0.7242461442947388)]
```

SCI-FI

HORROR

Nexts Steps and Caveats

Time X Genre

Find a way to extract specific common topics of discussion by characters

Movies belong to multiple genres

Extract dialogue(Potentially look into the Bechdel Test)

Other Gender Identities